

June 2013

## Pelikan's Antidisambiguation: "What's in a Name?"

Michael P. Pelikan

*Pennsylvania State University*, mpp10@psu.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Pelikan, Michael P. (2013) "Pelikan's Antidisambiguation: "What's in a Name?," *Against the Grain*: Vol. 25: Iss. 3, Article 42.  
DOI: <https://doi.org/10.7771/2380-176X.6549>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# 2013 Charleston Conference — 33rd Annual Issues in Book and Serial Acquisition

**Call For Papers, Ideas, Conference Themes, Panels, Debates, Diatribes, Speakers, Poster Sessions, Preconferences, etc. ...**

## 2013 Theme — Too Much Is Not Enough!

Wednesday, November 6, 2013 — Preconferences and Vendor Showcase

Thursday-Saturday, November 7-9, 2013 — Main Conference

Francis Marion Hotel, Courtyard Marriott Historic District, Charleston, SC

Addlestone Library, and School of Science and Mathematics Building, College of Charleston

If you are interested in leading a discussion, acting as a moderator, coordinating a lively lunch, or would like to make sure we discuss a particular topic, please let us know. The **Charleston Conference** prides itself on creativity, innovation, flexibility, and informality. If there is something you are interested in doing, please try it out on us. We'll probably love it...

The Conference Directors for the **2013 Charleston Conference** include — **Beth Bernhardt**, Principal Director (UNC-Greensboro) <beth\_bernhardt@uncg.edu>, **Glenda Alvin** <galvin@Tnstate.edu>, **Adam Chesler** <adam.chesler@cox.net>, **Cris Ferguson** (Furman University) <cris.ferguson@furman.edu>, **Rachel Fleming** (Western Carolina) <rflaming@email.wcu.edu>, **Joyce Dixon-Fyle** (DePauw University Libraries) <joyfyle@depauw.edu>, **Chuck Hamaker** <cahamake@email.uncc.edu>, **Tony Horava** (University of Ottawa) <thorava@uottawa.ca>, **Albert Joy** (University of Vermont) <albert.joy@uvm.edu>, **Ramune Kubilius** (Northwestern Health Sciences Library) <r-kubilius@northwestern.edu>, **Erin Luckett** (Readex) <eluckett@newsbank.com>, **Corrie Marsh** <cmarsh12@gmail.com>, **Jack Montgomery** (Western Kentucky University) <jack.montgomery@wku.edu>, **David Myers** <dave@dmmediaassoc.com>, **Audrey Powers** (UFS Tampa Library) <apowers@lib.usf.edu>, **Anthony Watkinson** (Consultant) <anthony.watkinson@btopenworld.com>, **Katina Strauch** (College of Charleston) <kstrauch@comcast.net>, or [www.katina.info/conference](http://www.katina.info/conference).

Send ideas by **July 8, 2013**, to any of the Conference Directors listed above.

**Or to:** Katina Strauch, MSC 98, The Citadel, Charleston, SC 29409 • 843-723-3536 (voice) • 843-805-7918 (fax) 843-509-2848 (cell) • <kstrauch@comcast.net> • <http://www.katina.info/conference> 🐼



## Pelikan's Antidisambiguation — "What's in a Name?"

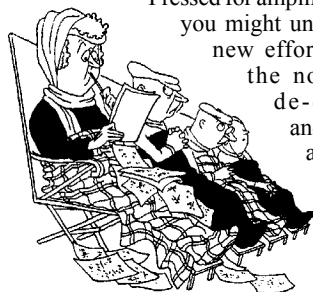
Column Editor: **Michael P. Pelikan** (Penn State) <mpp10@psu.edu>

**Juliet Capulet's** question has always carried weight with the communities that meet at *Against The Grain*. The answer to her question, I suppose, depends upon which community you happen to represent.

If you're a Librarian in the Technical Services, your answer may be, "A whole lot!" Pressed for amplification you might respond by revealing to the uninitiated the existence of the Authority System, Service, Database, etc. You might reveal the care and feeding, the accumulated person-centuries that have gone into the establishment and maintenance of a means to resolve issues concerning Name.

If you're a vendor of sufficient heft, your answer may be, "A huge potential market!"

Pressed for amplification (or not) you might unveil a massive new effort to monetize the normalization, de-duplication, and (now I actually must use the word) the disambiguation of Name information in connection



with authorship, especially in journals you vend — but probably, in your heart of hearts, Name information, well, maybe everywhere! If as a vendor, you're not thinking that big, rest assured, someone else is.

This seems an apropos time to pause and reflect on the present state of Name. A quick survey of my calendar shows just how many projects associated with some facet of Name are in motion. Here at **Penn State**, a major effort to rebuild the systems and methods for handling Name is coming on line after extensive effort and development. Meanwhile, the Internet2 community has been working on the issues associated with "consuming" social identity names at the institutional level. The ORCID project is gaining traction, and underscoring work already in place in professional and discipline-centric associations. And, the VIVO project continues to mature. I've just returned from having spent several delightful and enlightening days with the equally delightful and enlightened researchers, developers, and programmers at the heart of the VIVO project, and it's very healthy indeed. Let's touch on the items in this list one at a time.

**Penn State**, not too unlike many large universities, developed computer systems for

administering me-related information long ago. As it happened, we developed separate systems for handling Name, one for Students, the other for Faculty and Staff. There were yet other systems for prospective students, for alumni, etc. The list does go on, and at the scale that comes with a university such as **Penn State**, the numbers are impressive (we do call it the Big Ten, not the "Fairly Large Ten"). These systems were (and are) well and truly separate. Each had (has) its own representation for Name, for addresses, etc. A few years ago, a major (very major) effort was kicked off to bring all representations of Name (Person Names) at **Penn State** under one system, more or less (more rather than less). At the heart of the system is a new **Central Person Registry**, or **CPR**. In the course of an orderly transition, the **CPR** will become what we call the Authoritative Source for Person Name at the university.

Space does not permit me to delve deeply into the complexities involved with something so simple as Name, as represented in the context of a huge university. Although "e" issues are mind numbingly intricate at times, the ultimate goal of the project amounts to a

*continued on page 10*



# Take a closer look at....

## The CHARLESTON REPORT

*Business Insights into the Library Market*

### *You Need The Charleston Report...*

*if you are a publisher, vendor, product developer, merchandiser, consultant or wholesaler who is interested in improving and/or expanding your position in the U.S. library market.*

*Subscribe today at our discounted rate of only \$75.<sup>00</sup>*



**The Charleston Company**

6180 East Warren Avenue, Denver, CO 80222

Phone: 303-282-9706 • Fax: 303-282-9743

### **Pelikan's Antidisambiguation** from page 8

vast simplification, clearing away sources of confusion and thrashing that result from having many separate "silos" in which something so fundamental as Name is stored. The savings and resulting efficiencies to be gained over the life of the system will recoup the investment of time and treasure in development — and then some.

As for what we call Social Identifiers — the Internet2 community has been delving into what it would take and when it would be appropriate to handle the identities students, parents, and others come to our institutions with — identities they already possess, that they got from having accounts with Google, Microsoft, Facebook, etc. There's philosophical consensus that for low-stakes kind of interactions, folks ought to be able to interact with our services with the identities they already have — that we needn't insist they create a new university-hosted account merely to get information about a program, for example. When things become interesting is the point at which the relationship gets taken to the "next level," and we begin to need greater assurance that the person we're dealing with is really whom he or she asserts themselves to be, as well as being the same person who asserted that identity the last time we saw it. The opportunities for standards and exchangeable, sharable information drive us to discussion and coordination of approach.

I should mention work underway with the **Educational Testing Service** to pass forward a high-quality, vetted identity with the records associated with students taking SAT tests. Potentially, the very first identity we as a university receive regarding a prospective student could be one that has been carefully proofed, as well as bound, to a physical individual who arrived with identity credentials required to take the SAT test. If this effort works out, it will be a huge win for colleges and universities everywhere, and will mitigate the "merging and matching" that has to go on behind the scenes as prospects with a Gmail account become "paid accepts" with a university ID.

The ORCID project (<http://orcid.org>) is an effort to provide "a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkage between you and your professional activities..." That's a direct lift from the ORCID site which, if you haven't visited yet, would be a worthwhile browse. It's true that within disciplines, professional or author identifiers furnished through professional organizations such as **ACM** or **IEEE** provide something of this. Indeed, there is no conflict, from an information science perspective, between such efforts. When it comes to (grumble) "disambiguating" an identity, in some respects, the more attributes we can get, the better — so long as they've been applied with care and some certain degree of rigor.

In many respects, it's in work such as the VIVO community has undertaken that all these

efforts come together semantically — literally! It is in the representations of relatedness available through RDF (<http://www.w3.org/RDF/>), the Resource Description Framework, that all these many and varying bits of information about persons, the names they use, the identifiers they've accumulated, the efforts they've been involved in, the institutions they've been associated with and the roles they've played, the projects they've worked on and the works they've published — all of this can be tied together in a massively huge, massively diverse, massively consistent representation — truly, the Semantic Web realized.

Of course, RDF is an open standard, and VIVO is an open source project. Among the first to recognize its potential have been, unsurprisingly, the vendors who publish the products of research and sell access to those products back to the universities. The very large vendors have both the scale of perspective and the deep pockets needed both to support and ultimately, to profit from, the kind of opportunity project and products such as VIVO, **Harvard Profiles**, and **Digital Measures Activity Insight**. At the heart of these efforts, besides open standards such as RDF, there are ontologies. And ontologies, at an intercontinental scale, represent a vast frontier without fences, ripe, fertile, and ready to be claimed and staked. It's admittedly complicated stuff. All the better! Turn the underlying enabling technology into a product that can be subscribed to, make it cheap (a relative term) enough not to kill the customer, yet expensive enough to require high-level negotiation and approval, and you can effectively wrest control of the effort away from the scary Semantic Web Eggheads at the institutions and turn it into respectable, forward-leaning suite of products and services from reputable vendors with global reach. Oh, and you can make it simple, too. None of that mind-numbing complexity. "We already have the information you need," the pitch will go, "... just sign here."

I fully appreciate the capabilities of the large vendors to support, to buttress, the underlying information environment upon which all such efforts rely. I also stipulate, up front, that they have what it takes to make high quality products and services in this space. But before you sign, please consider what the ramifications would have been if the **AACR** or **Library of Congress Subject Headings** had been born out of any motivation beyond merely the Public Good.

We're at a time about which an extended and wholly appropriate analog can be drawn to another frontier time: the conceiving, lobbying, financing, and building of the first Transcontinental Railroad and all which that entailed. We have the visionaries, the technicians, the promise of new and previously unattainable connections, the pathway for prospectors and homesteaders, the ushering in of a new age.

"The Semantic Web, realized" is the shape and substance of the coming information age. There's clearly enough to go around for everyone, eggheads AND vendors, to collaborate, cooperate, and work on together. 🍄